MODIS DATA SYSTEM STUDY

TEAM PRESENTATION

October 7, 1988

AGENDA

1. MIDACS Element Functional Allocation Diagram

2. Quick-Look Data Requirements of the MODIS Instrument Team

   -- Instrument Monitoring
   -- Support of Field Experiments

3. MIDACS DBMS Overhead Considerations

4. CDHF 4.3 Data Flow Diagram

5. Action Items

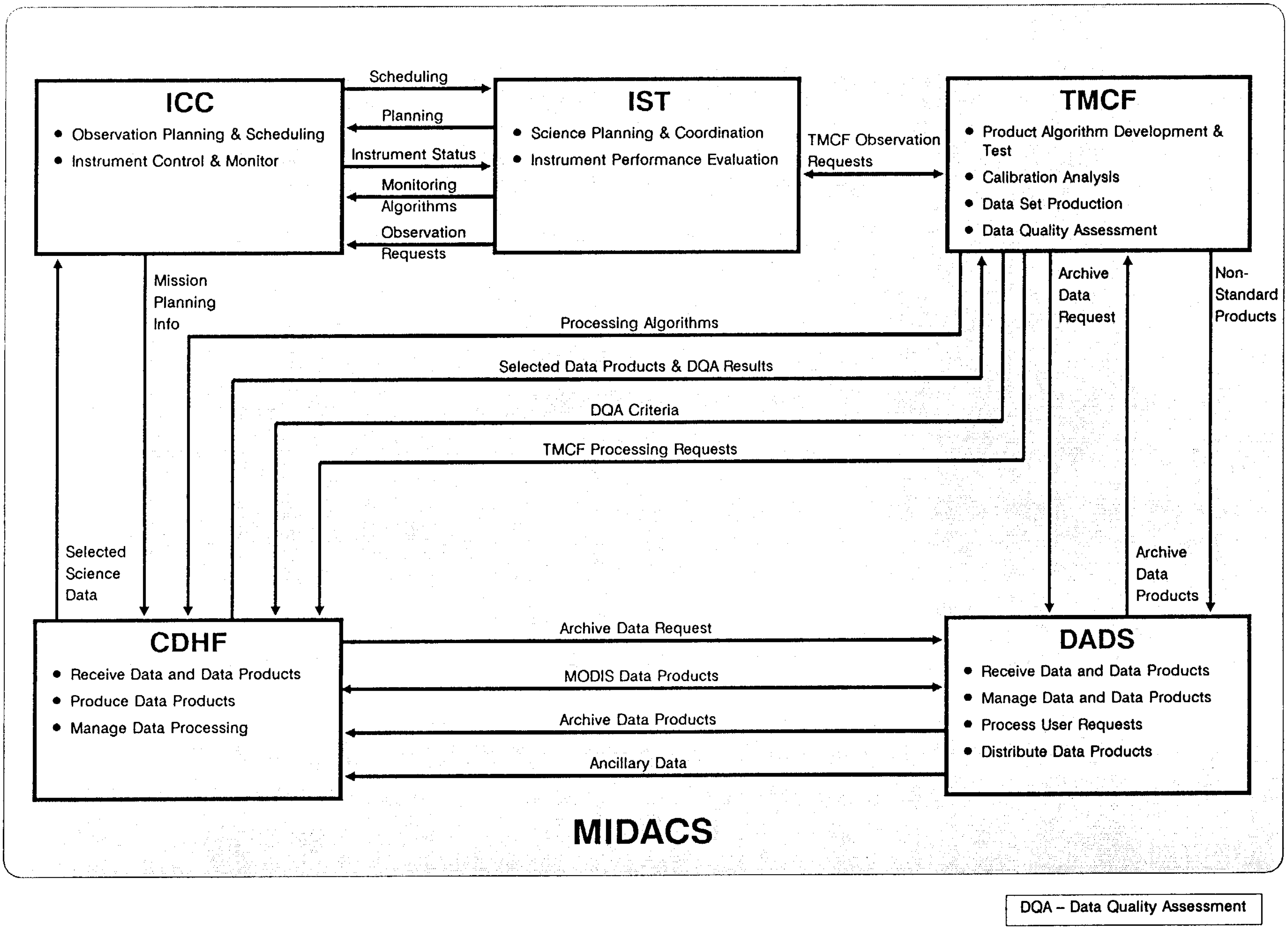


FIGURE 2. MIDACS Element Functional Allocation Diagram

QUICK-LOOK DATA REQUIREMENTS OF THE MODIS INSTRUMENT TEAM

For the purposes of discussion here, we define "quick-look" data as MODIS standard and specialized (non-standard) data products generated more rapidly than normally associated with standard processing. This quick-look data is subdivided, in turn, into "real-time" and "near-real-time" data.

We define real-time data, and the associated "real-time data processing" within MIDACS, as data for which no delays in processing have been incurred (e.g., transmission path delayed only). In other words, from the time the MIDACS receives the MODIS data, the output products are created without any buffering or storage delays. We define near-real-time data, and the associated "near-real-time" processing within MIDACS, as data for which delays have been incurred, though not to the extent of routine processing.

We have received quick-look data requirements for both real-time and near-real-time science data. The real-time data is required to monitor the MODIS instrument performance in real-time. The near-real-time data is required to support atmospheric, oceanographic, and land science field experiments.

MIDACS NEAR-REAL-TIME REQUIREMENTS

1. The MODIS data system shall support atmospheric, oceanographic, and land science field experiments by providing near-real-time processing of MODIS data.

2. All near-real-time data processing shall be completed within three to eight hours at all levels (Level-1A to Level-4). The precise timeliness requirements will depend on the specific data requirements of the field experiment.

3. The data system shall be sized to support 15 simultaneous field experiments each for MODIS-N and MODIS-T: five for atmospheric sciences, five for oceanography, and five for land-sciences (a total of 30).

4. For each field experiment supported, a set of scenes shall be generated. A scene is defined as having the dimensions of a full cross-track scan width (i.e., 1294 pixels or about 2,000 km) by 2,000 kilometers (i.e., about 2,048 one-kilometer detector swaths or 5 minutes of data). A set of scenes shall include the calibrated and Earth-located (Level-1B) radiances for 15 to 20 spectral channels, 25 Level-2 parameters, and TBD Level-3 and Level-4 products.

5. Communications capabilities shall be imbedded into the data system so as to enable the delivery of near-real-time scene data to the investigators at the site of the experiment within the specified timeliness requirements.

6. Computing resources shall be made available to the investigators at the site of the experiment (perhaps as portions of the distributed TMCF) to enable in situ analysis of the MODIS data products.

MIDACS REAL-TIME REQUIREMENTS

1. To facilitate the analysis of real-time data, the MODIS data shall be packetized on board the platform by channel, detector, and scan line (1,292 pixels by 12 bits corresponds to 1,938 8-bit bytes).

2. The MODIS data shall be buffered on board the platform for a complete scan to permit the required packetization of the data.

3. On-board storage of 8 megabytes for MODIS-T (64x64 x 1,294 x 12/8) and 1.5 megabytes for MODIS-N (752 x 1,294 x 12/8; 752 = 30 x 8 + 8 x 32 + 2 x 128) shall be required to provide on-board buffering capabilities.

4. The ICC/IST shall provide the MODIS Instrument Scientist and/or his designated representatives the facilities to monitor the MODIS instrument performance by making available all the science instrument data in real-time.

5. Facilities for real-time MODIS instrument monitoring shall include two interactive workstations; one each for the MODIS-N and MODIS-T instruments.

5. The scientist monitoring the real-time MODIS data shall be able to select, randomly or systematically, any four MODIS channels for simultaneous analysis for MODIS-N and for MODIS-T (a total of eight channels).

6. Upon selection, data from the designated channels shall begin to build 2000 km x 2000 km scenes in the workstations' memory <u>without delay</u>. Once built, the scenes shall be continuously refreshed.

7. The DHC shall supply to the ICC, and the ICC shall be designed to accept, the entire MODIS instrument data stream in either real-time or priority-playback mode [derived from requirements 5 and 6].

8. Each workstation shall have sufficient internal RAM to store and manipulate the four scenes simultaneously (2,048 x 2,048 x 4 x 2 bytes --> more than 34 megabytes), as well as 200 megabytes of on-line storage, TBD off-line storage, and TBD hard-copy output devices.

9. Each workstation shall be capable of performing TBD analysis of the data.

10. The workstations shall be capable of simultaneously performing both the display and analysis functions.

The following questions are directed towards the capability of the DHC to fulfill the MIDACS requirement.

1. Can the DHC send instrument packets in real-time to cover the concept mentioned above?

2. Can the DHC provide priority playback data to the MIDACS for this requirement? How would the DHC identify the requested data?

3. Can the DHC provide both real-time and priority playback data to the MIDACS simultaneously?

4. Is the availability of real-time data dependent on the requirements of other instrument facilities? If so, how should the MIDACS request real-time or priority playback data to satisfy the above requirement?

5. What is the average and peak output data rate of real-time or priority playback data from the DHC to the MIDACS?

DBMS OVERHEAD

Relevance and Motivation

DBMS overhead refers to additional storage or processing necessary solely for supporting the DBMS package. This overhead has both fixed and variable components, both being partially amenable to fine tuning. Predicting DBMS overhead levels is often a complex task as many variables enter into the calculations. Based on application attributes, levels from 10 percent to 100 percent can be realized.

As one aspect of our DADS Operational Concepts Report, we considered the long-term MODIS data volumes. These estimates incorporated an overhead of 20% for cataloging and management. One key element of this data system study is the sizing of the system. Clearly, the sizing of the DBMS in terms of both storage and processing overhead are critical elements; these will be addressed in depth as this study progresses.

Storage Overhead

DBMS storage overhead refers to the additional storage a DBMS uses, above and beyond what is necessary for storing data values. The DBMS' data model, vendor preferences, and user application attributes contribute to the overhead level. Each DBMS implements a logical data model, three main models being hierarchical, network, and relational. As the relational model is the most popular model, MODIS could be expected to use it. The vendor's implementation via the physical model impacts overhead level. Additionally, vendors may choose to optimize their packages for specific activities such as update or query.

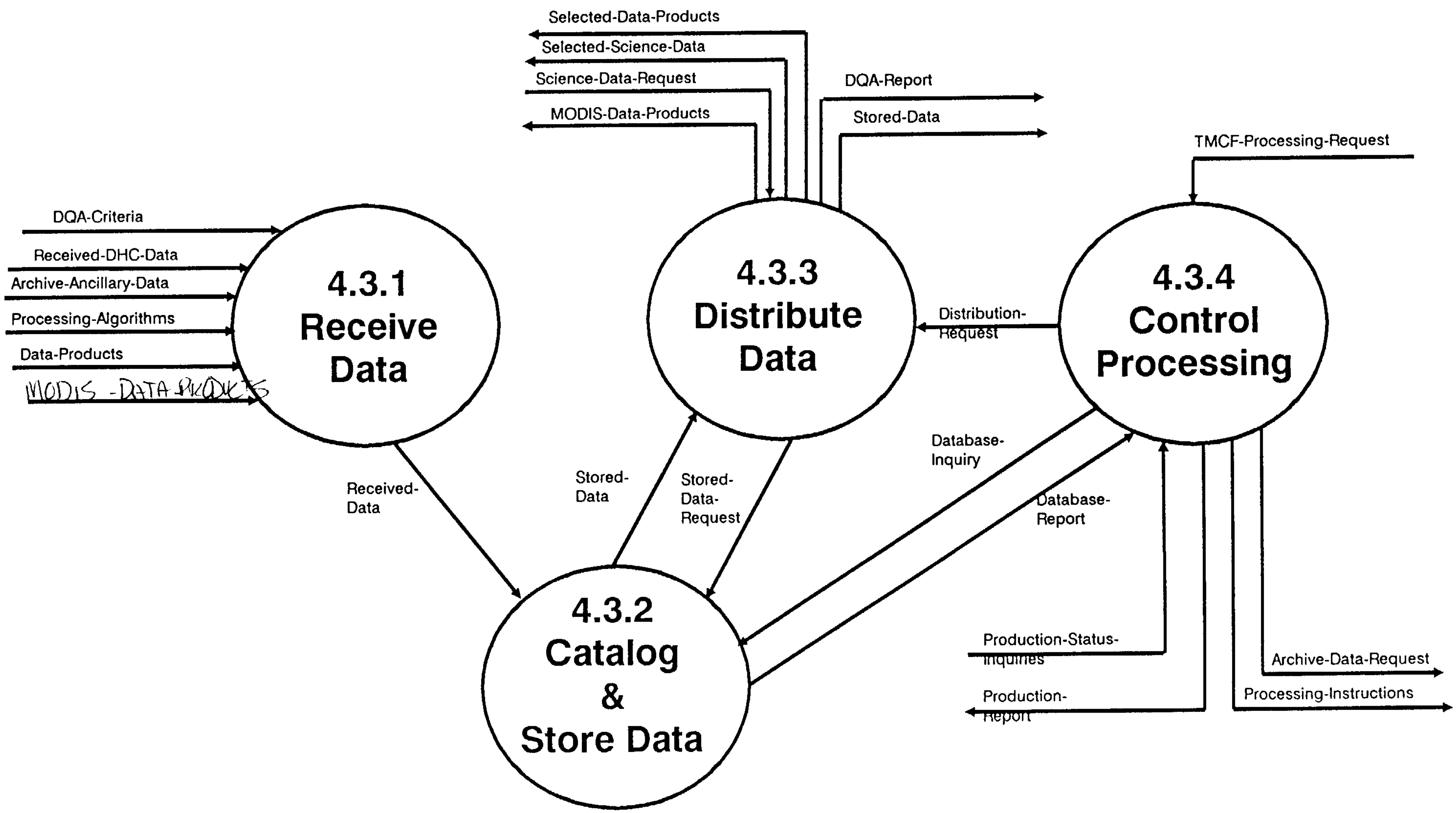
Each application's nature and intended users also influence the overhead rate. Data elements can be defined as key or non-key. Key definitions require pointers to similar values possessed by other occurrences of that data element. The more varied a data element's values, the more overhead. Each data element requires a separate data dictionary entry. These entries can describe value ranges, access restrictions, synonyms, and other attributes. The more data elements, the more overhead. Complex dictionary entries also contribute to overhead.

Databases designed to support browsing and ad-hoc queries tend to have a higher level of key items, hence higher levels of overhead. This is necessary to support queries with substantial numbers of Boolean operators that are used to qualify on large numbers of data elements and data element values. Databases in less dynamic environments would require fewer key elements, meaning lower overhead.

Processing Overhead

Processing overhead represents the DBMS' internal processing to parse queries, determine permissable accesses via the dictionary, and formulate the actual database accesses necessary for meeting a user request. As DBMS' are generalized packages, some part of this processing overhead may be unnecessary in a given application. Individual vendors differ in the degrees of fine-tuning their packages offer.

Processing overhead is also a function of distributed processing. Its level is in part determined by the definition of "distributed." The term can refer to multiple copies of a single database, partial copies of this same database, or single copies of database subsets that when combined represent the entire database. The term can also refer to a single DBMS copy or multiple DBMS copies in conjunction with a "distributed" database. DBMS' differ in the degrees of efficiency with which each of these configurations can be implemented.



CDHF DFD 4.3 Control Processing & Handle Data

ACTION ITEMS:

9/16-1: (McKay) Does MIDACS need to request Level 0 data from the DHC routinely, or will the DHC send Level 0 data to MIDACS as soon as it is ready without an explicit request?

9/30-1: (Han) Provide a list of candidates to send the MODIS Data System Questionnaires to.

9/30-2: (Ardanuy) Define conceptually how the science team members will get their data (assume that the TMCF is distributed and that 9.6 kbps is not a sufficiently high data rate).

9/30-3: (Folta) Under what conditions will the DADS release data? Will the data be released automatically unless there is a notification to hold the data or will the data only be released upon explicit authorization of the science team?